

THE COUNCIL FOR TOBACCO RESEARCH-U.S.A., INC.

110 EAST 59TH STREET
NEW YORK, N. Y. 10022

July 25, 1975

Messrs. Thomas F. Ahrensfield
DeBaun Bryant
Cyril F. Hetsko
James Scott Hill
H. C. Roemer

Gentlemen:

A review has been made of the operations of the Literature Retrieval Division. This covers the methods of handling the literature and also the computer process. The former was prepared jointly by Dr. Colby and Dr. Giller, the latter by William Benbow of Reynolds. For your information copies are enclosed.

Sincerely,

H. H. Ramon
H.H.

H. H. Ramon
President

HHR:ek

encls.

bcc: FBG, WTH, HHR, AY

Literature Retrieval Division Operations

July 22, 1975

A. OBSERVATIONS

I. Quality of Document Retrieval

With regard to subject searches, the overall conclusion is that the System operates satisfactorily because in-house System professionals have been trained and made available to provide interfacing and because there has been sufficient turn-around time available for search formulation and evaluation of search results. ("Satisfactory" is defined here as the ability to retrieve most of the on-target articles stored in the data bank and to provide these to Users for detailed analysis.)

Search Formulation

As is characteristic of any automated information storage and retrieval system, special skills are required to formulate search questions--at times through trial and error. Very few Users ever have expressed interest in learning the required special skills. In addition, a substantial number of documents processed during the earlier days of the System are not indexed according to current standards, and familiarity with past and present indexing standards is required to ensure that retrospective searches are formulated properly.

Evaluation of Search Results

Search results are reported as either a list of accession numbers of documents or as a "Full Report" of all information stored in the data bank for each document.

When search results are reported by accession number only and there is a large number of "hits", considerable time must be spent pulling the aperture cards and reviewing the abstracts and index terms in order to identify the on-target articles. At this preliminary stage of evaluation, identification of on-target articles consists largely of excluding articles which do not appear to be on-target. This is consistent with the classical use of any abstract; in our System as well, the abstract was never intended to be a substitute for the original article.

In later stages of evaluation of search results, it is often necessary to peruse the text of the original article. This is because a substantial number of articles processed during the earlier days of the System are not abstracted

according to current standards, the exclusionary potential of an abstract varies with the goal of each specific search, and, as already noted, the abstract was never intended to substitute for the original article. Thus, the text of the article often must be perused for both currently produced documents and documents produced earlier--much less frequently for those produced currently.

When search results are reported by a Full Report of all information stored in the data bank for each document, the preliminary evaluation procedure is facilitated because certain documents can be excluded by a review of the data displayed. However, the abstract is not stored in the data bank, and therefore, does not appear in the Full Report display, and the print format and sorting options of the Full Report System have been cumbersome and limited.

A new computer system, which has been in the development and testing stage for approximately five months, recently has been implemented to eliminate the cumbersome and limiting aspects of the Full Report display. The sorting options of the new system have been expanded to their maximum possibilities and a variety of print formats are available. This new system has been operational for only a few weeks, but initial indications are that total search evaluation time will be reduced by approximately 20 to 30 percent. (The Literature Retrieval Division will be publishing an addendum to the User Manual describing the uses and capabilities of the new system; a brief description is attached here--refer to Attachment A.)

Even with the implementation of the improved Full Report display, the final stages of search evaluation require pulling aperture cards, reviewing abstracts and index terms, and in a number of cases, the text of the article. In addition, if the results of an initial search do not retrieve the desired documents, a follow-up search may be required, although both broad and narrow searches routinely are run simultaneously to avoid such situations. Follow-up searches obviously add to the total turn-around time.

Thus, in a case of an "emergency" (i.e., when the available turn-around time is less than four to eight hours and there is a large number of "hits" and computer search time and additional qualified back-up personnel are unavailable) it is possible that many on-target articles will not be retrieved; however, some on-target articles always will be retrieved even in the event of such an extraordinary combination of circumstances.

II. Quality of Abstracting and Indexing

The quality of the abstracting and indexing has improved substantially since the "3i and O'Shea eras," especially since the abstracts have been structured.

There is no question that the quality of the abstracting and indexing is much better than that provided by any information service in the public domain and that to bring the quality to a "perfect" level would involve an unproductive amount of additional intellectual effort.

With regard to abstracting, one of us (F.C.) believes that "many" of the abstracts are unnecessarily verbose and that in these cases all information in the abstract might just as well be provided with "half or one third less words;" as a corollary, additional pertinent information could be included using the same number of words. The other of us (F.B.G.) is deferring judgment on this observation, subject to a quantitative study of the question.

With regard to indexing, no major deficiencies were detected in a sample of documents recently evaluated by F.C. However, in F.C.'s judgment, there were still imperfections in some of the documents.

III. Quality of Scanning and Selection

The quality of scanning and selection is completely satisfactory.

IV. Document Backlog

There exists a significant backlog of unprocessed documents even though the Literature Retrieval Division has increased productivity substantially over the years. This backlog is due to at least three major factors: 1) more in-Scope articles are appearing in the published literature. (The prior consensus of all individuals involved with the System that the number of in-Scope articles probably would diminish with time has proved to be incorrect.); 2) the definition of an in-Scope article has been expanded considerably; 3) the scanning efforts of the Literature Retrieval Division have improved substantially.

Because of the backlog, a major goal has been to ensure that "high priority" documents are processed preferentially. Reviews of the backlog and selection of documents for preferential processing are performed routinely by Literature Retrieval Division personnel in accordance with changing User priorities.

However, based on a study of a sample of documents not yet shipped to Users, a few priority items were identified. As a result, F.C. reviewed the "To Be Abstracted" backlog of articles published during 1970 through 1974. Ten priority documents (six of which are foreign language) and 229 other documents (85 of which are foreign language) were selected by F.C. for preferential processing.

V. EDP Personnel

The quality of the in-house personnel needed to operate and maintain the computer system is completely satisfactory.

B. RECOMMENDATIONS

I. Document Retrieval

Several recommendations which require further study have been discussed for improving turn-around time and other facets of document retrieval. Since the quality of document retrieval is satisfactory, the implementation of these recommendations is not an urgent matter.

These recommendations are listed in order of decreasing priority.

- *Automated retrieval of abstracts for use during preliminary evaluation of search results without the necessity of pulling aperture cards.

- *Weighted citation indexing.

- *The possibility of an On-Line System with visual display.

II. Abstracting and Indexing

F. C. should continue to monitor the output of the System and to discuss any suggested changes on an article-by-article basis during one-day meetings in New York City about six times per year, or more frequently if needed.

III. Scanning and Selection

F. C. should continue to make available on a regular basis to the System the results of his scanning and selection from approximately 20 primary journals and a few secondary sources. Differences of opinion regarding input of specific articles should be discussed during one-day meetings in New York City about six times per year, or more frequently if needed.

IV. Document Backlog

All English language articles published during 1970 through 1974 that were selected for preferential processing will be input into the data bank on or before November 1, 1975 (approximately 150 articles); all such foreign language articles will be input on or before April 1, 1976 (approximately 90 articles).

Thereafter, it is not anticipated that a backlog will develop of "current" articles (i.e., articles published during 1973 or later). Thus, the "backlog" consists essentially of low priority articles published during 1972 and earlier--refer to Attachment B.

The time that will be required to process the backlog will depend on the revised Scope of Coverage to be adopted and the training of additional abstractors, editors and supporting personnel.

Attachment A

DESCRIPTION OF THE NEW FULL REPORT SYSTEM

The User has the options to specify the print format, the parameter for sorting, and the parameter(s) for display.

Print Format (select one):

Computer paper compactly in 8 1/2 x 11 inch borders
 Tab (IBM) card, one reference per card
 3 x 5 inch card, one reference per card

Sorting Parameters (select one):

Category Number(s)
^aPrimary Author only
 Secondary Author(s) only
^aPrimary and Secondary Authors
 Title (first 64 characters)
 Journal
 Year
 Primary Affiliation
 Descriptors

Display Parameters (select one or more):

Accession Number
 Category Number(s)
^bPrimary Author only
^bPrimary and Secondary Authors
 Title
 Journal, Volume (issue), page
 Year
 Affiliation(s)
 Descriptors
 Weight
 Sort Parameter

Notes:

- ^aSorting will occur on Tertiary Author(s) if Primary Author is Anonymous and a display parameter is Primary Author.
^bTertiary Author(s) will be reported if Primary Author is Anonymous.

Attachment B

DOCUMENT BACKLOG

Documents in Process (June 30, 1975 Production Report)

Total in process	2,803
Documents loaded, but not shipped	<u>- 273</u>
Total in process adjusted for lag in aperture card shipments	<u><u>2,530</u></u>

Documents in Process--Revised

	<u>Year of Publication</u>	
Working backlog ("current")	1973-1975	950 ^a
Selected for preferential processing by F.C.	1970-1974	240
"Backlog" of low priority articles	1972 & earlier	<u>1,340</u>
		<u>2,530</u>
Rejected by F.C. (borderline articles)	1970-1974	<u>- 213</u>
Revised total in process		<u><u>2,217</u></u>

Note:

^aA working backlog is necessary to ensure a continuous production flow. The present working backlog consists of enough documents for approximately three months' output.